An R package facilitating sensitivity analysis, calibration and forward simulations with the LPJ-GUESS dynamic vegetation model

Maurizio Bagnara[a,b,*], Ramiro Silveyra Gonzalez[a,c], Stefan Reifenberg[d], Jörg Steinkamp[b], Thomas Hickler[b,e], Christian Werner[b], Carsten F. Dormann[a], Florian Hartig[a,f]

[a] *Department of Biometry and Environmental System Analysis, University of Freiburg, Tennenbacher Straße 4, 79106, Freiburg im Breisgau, Germany*
[b] *Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325, Frankfurt am Main, Germany*
[c] *Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, P.O. Box 601203, D-14412, Potsdam, Germany*
[d] *Chair of Forestry Economics and Forest Planning, University of Freiburg, Tennenbacher Straße 4, 79106, Freiburg im Breisgau, Germany*
[e] *Department of Physical Geography, Geosciences, Goethe-University, Altenhöferallee 1, 60438, Frankfurt am Main, Germany*
[f] *Theoretical Ecology, Faculty of Biology and Preclinical Medicine, University of Regensburg, Universitätsstraße 3, 93053, Regensburg, Germany*

ARTICLE INFO

ABSTRACT

Dynamic global vegetation models (DGVMs) are of crucial importance for understanding and predicting vegetation, carbon, nitrogen and water dynamics of ecosystems in response to climate change. Their complexity, however, creates challenges for model analysis and data integration. A solution is to interface DGVMs with established statistical computing environments. Here we introduce rLPJGUESS, an R-package that couples the widely used DGVM LPJ-GUESS with the R environment for statistical computing, making existing R-packages and functions readily available to perform complex analyses with this model.

We demonstrate the advantages of this framework by using rLPJGUESS to perform several otherwise laborious tasks: first, a set of single simulations, followed by global and local sensitivity analyses, a Bayesian calibration with a Markov-Chain Monte Carlo (MCMC) algorithm, and a predictive simulation with multiple climate scenarios. Our example highlights the opportunities of interfacing existing models in earth and environmental sciences with state-of-the-art computing environments such as R.

Data availability

Climatic data used in the simulations have been obtained from the Biodiversity Exploratories Database (https://www.bexis.uni-jena.de).

The data used for the simulations are available at https://drive.google.com/uc?id=1JE7H0duYxoWfxYwb_kVUxOLrfOujaNx0&export=download.

Software availability

The rLPJGUESS source code is publically available at https://github.com/biometry/rLPJGUESS. rLPJGUESS requires R v. >= 3.1.0, and depends on R-packages snow and zoo, both available on CRAN. rLPJGUESS requires a compiled version of LPJ-GUESS v.3.1 (Smith et al., 2014), the source code of which is available upon request at http://web.nateko.lu.se/lpj-guess. NOTE: rLPJGUESS is not compatible with the Educational Version of LPJ-GUESS publically available on the LPJ-GUESS website.

1. Introduction

The complexity of terrestrial ecosystems, with many interactions between the various drivers and compartments, limits our ability to understand and predict system responses based on simple statistical relationships. Therefore, the most widely used tools to describe transient vegetation dynamics in response to changes in climate and atmospheric CO_2 concentration are dynamic, process-based vegetation models (Hartig et al., 2012). These models include detailed process-based representations of productivity, competition, population dynamics and forest succession, either for particular tree species (e.g. Bugmann, 2001), or in the so-called dynamic (global) vegetation models (D(G)VMs) usually for plant functional types (PFTs e.g. Prentice et al., 2007).

Most D(G)VMs are fairly complex (Dietze et al., 2014). Working with such complex system models creates distinct challenges:

1. Simulations are often slow, necessitating the use of parallel

---

Table 1
Typical tasks for model analysis, and general estimates of the number of model runs required for their calculation. In practice, the actual number of runs may depend on several factors such as model characteristics, algorithm used, or the number of parameters under consideration (Bagnara et al., 2018).

Goal/Aim		Method	Number of model evaluations	Parallelizable
Model evaluation	Calculate scenarios	Run model with different parameters/inputs	1–10.000	Yes
Sensitivity Analysis	Vary inputs/parameters to learn about model reaction	Many local and global methods	5.000–100.000	Yes
Inverse calibration	Vary parameters to find the best fit	Optimization algorithms	1.000–50.000	Partly
		MCMC	50.000–10 mio	Partly
Monte-Carlo Uncertainty Analysis	Forward uncertainties to outputs	MC draws from all sources of uncertainty	1.000–10.000	Yes

computing when many model evaluations are necessary.

2. The complexity of process interactions often leads to unpredictable model behaviour from its structural properties – simulations need to be run to understand how the model responds to changes in inputs or parameters. Algorithms to perform such local or global sensitivity analysis (Saltelli et al., 2000) often require a considerable number of model evaluations (Table 1).

3. Even more evaluations are needed for model inversion (Table 1). Model inversion, or inverse calibration, means that we search for the model parameters and inputs most likely to produce the observed outputs (Voosen, 2016). In recent years, Bayesian approaches to model calibration have become increasingly popular, because they allow quantifying model parameters and predictions as well as their uncertainties (Bagnara et al., 2015; Hartig et al., 2012; van Oijen et al., 2005). As posterior densities usually cannot be derived analytically, Monte-Carlo algorithms such as importance sampling, Markov-Chain Monte Carlo (MCMC) and sequential Monte Carlo are commonly used (Hartig et al., 2011).

4. Finally, decision makers demand model predictions, together with their uncertainties. This is often achieved via Monte-Carlo simulations or ensemble runs. In this context, Bayesian methods provide a particularly comprehensive, if computer-intensive, way to propagate uncertainty in different parts of the model to the output.

Although all these techniques are well known and readily available in scientific computing environment such as R, MATLAB and Python, they are still rarely applied for complex DGVMs (e.g. Luo et al., 2009). In our experience, an important reason is the amount of work associated with interfacing established DGVMs, which are usually programmed in lower-level languages such as C/C++ or Fortran, with existing statistical frameworks and algorithms that are available in the mentioned higher-level languages.

For the LPJ-GUESS model (Lund-Potsdam-Jena General Ecosystem Simulator, Smith et al., 2014), we address this problem through a new interface with the statistical tools provided in R (R Core Team, 2015). LPJ-GUESS is a process-based model of vegetation dynamics and biogeochemistry designed for regional to global applications. It combines eco-physiological features of the widely-used Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ-DGVM; Sitch et al., 2003) with detailed representations of vegetation dynamics and canopy structure as used in forest gap models. LPJ-GUESS, developed by a community of scientists from different institutions (http://iis4.nateko.lu.se/lpj-guess/index.html) is one of the most widely applied DGVMs to investigate vegetation dynamics, so far resulting in more than 200 international peer-reviewed articles in ISI-listed journals (http://iis4.nateko.lu.se/lpj-guess/LPJ-GUESS_bibliography.pdf). The model has also been parameterized regionally (e.g. its "European" version represents main tree species instead of global PFTs, Hickler et al., 2012). Recently, Smith et al. (2014) incorporated nitrogen cycling and limitation, making LPJ-GUESS one of the most complete vegetation models currently available.

Despite, or maybe because of this complexity, however, most applications of LPJ-GUESS concentrate on predicting or understanding present and past patterns and dynamics (Ahlström et al., 2015), while few studies present advanced analysis of the model itself, such as sensitivity analyses or model calibration (but see Pappas et al., 2013; Wramneby et al., 2008; Zaehle et al., 2005). In this paper, we present rLPJGUESS, a newly developed R-package that couples LPJ-GUESS with R: this more convenient interface to the many methods available in R will create new research possibilities for vegetation modellers, and thereby help to advance the field's capabilities to run complicated analyses and tests.

2. The rLPJGUESS package

With rLPJGUESS, every step necessary to run LPJ-GUESS can be
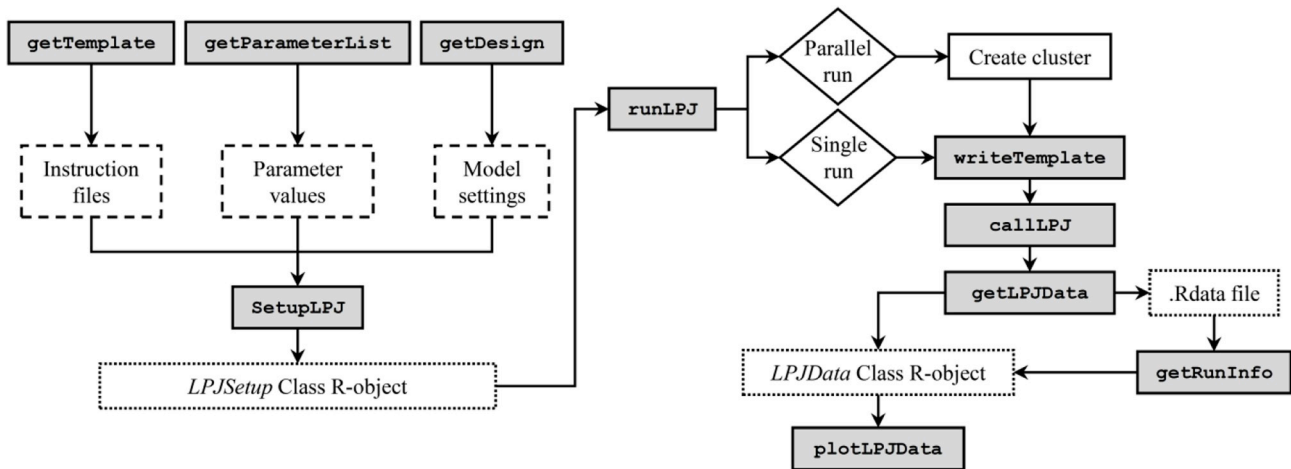
Fig. 1. `rLPJGUESS` workflow. The `runLPJ()` function is the core of the package, with functions on its left referring to model setup, and functions on its right relating to collating of output. Grey boxes represent user-available functions. Dashed boxes represent material provided within the package for model setup. Dotted boxes represent outputs from the previous function.

carried out from within R, facilitating model setup, output analysis and complex tasks such as sensitivity analysis and model calibrations, for which R provides many dedicated packages.

A basic use example of `rLPJGUESS` would be to run LPJ-GUESS for a single site with the default parameter values. `rLPJGUESS` allows to run LPJ-GUESS both in PFT or species-specific mode, and it supports the default climate data supplied by the LPJ-GUESS developers as well as user-defined climate input files. The standard run settings of LPJ-GUESS (e.g. presence/absence/length of the spin-up, species or PFTs simulated, nitrogen limitation etc.) can be changed through the `rLPJGUESS` functions `getDesign()` and `getParameterList()` (Fig. 1). After the model run has terminated, user-selected outputs are returned for further analyses, together with the settings of the model run (Fig. 1).

This basic use can be extended by:

1. changing model parameters, for example to carry out calibrations and sensitivity analysis; and
2. running the model on multiple locations, or with multiple climate scenarios.

For both options, `rLPJGUESS` supports parallelization for personal computers, servers and High-Performance Computing (HPC) clusters, including the Message Passing Interface (MPI) protocols for parallel computing typically available on HPC clusters. `rLPJGUESS` is fully operational on macOS, Windows and Linux operating systems.

3. Case study

To demonstrate the functionality of `rLPJGUESS`, we present a case study comprised of four tasks, carried out from within R:

1. a set of single model runs;
2. a global and local sensitivity analysis (using parallelization);
3. a Bayesian model calibration;
4. a parallel prediction run using multiple climate scenarios.

Installation instructions and code for every task are provided as Appendices to this paper.

3.1. Single model runs

To demonstrate the basic communication of the `rLPJGUESS` package with LPJ-GUESS, we performed three single model runs (Appendix B), using 62 years of weather data (1950–2012) obtained from the German Weather Service's station in Erfurt, Germany. For all the simulations, we estimated the vegetation biomass carbon (C-biomass) and the leaf area index (LAI), with the settings given in Table 2.

In the first run (Fig. 2a), we simulated vegetation dynamics with the PFT parameterization for a single forest stand, including only temperate PFTs that would be able to grow at the study site according to their bioclimatic limits (Smith et al., 2014). In the second run (Fig. 2b), we simulated the same stand with the species-specific parameterization including only temperate species (Hickler et al., 2012). In the third run, we simulated only *Fagus sylvatica* (Fig. 2c), which is the dominant species in the area (Fischer et al., 2010). The results are generally in line with expectations about the natural regeneration of a pure beech stand in a similar climate (4–7 Kg C m^{-2}, Granier et al., 2008). The calculation time was ∼2.5 s for each simulation.

Table 2
Simulation settings that were modified from the LPJ-GUESS default values following Hickler et al. (2012), simplifying the model setup for the purpose of this study.

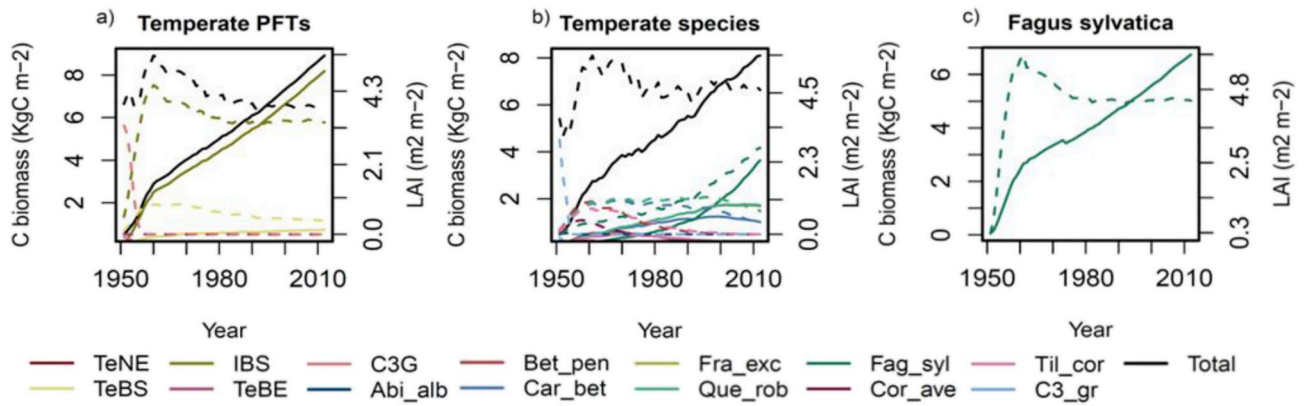| Setting	Value	Description
ifcentury	0	use CENTURY SOM dynamics (1) or not (0)
ifnlim	0	plant growth limited by available N (1) or not (0)
iffire	0	implement fire (1) or not (0)
ifstochestab	0	establishment stochastic (1) or not (0)
ifstochmort	0	mortality stochastic (1) or not (0)
patcharea	25^2	patch area (m^2)
npatch	1	number of replicate patches to simulate
ifdisturb	0	generic patch-destroying disturbances enabled (1) or not (0)
nyear_spinup	1	number of years to spin up the simulation for
freenyears	0	number of years to spin up without N limitation

Fig. 2. Simulations of total carbon (solid line) and leaf area index (dashed) of model runs for a) temperate PFTs, b) temperate species, and c) monospecific *F. sylvatica* stand. Abbreviations follow the LPJ-GUESS standard (see Appendix B, Fig. B1).

3.2. Parallelized sensitivity analyses

For the next analysis step, we used the simulated LAI and C-biomass for *F. sylvatica* as reference data, and created a goodness-of-fit function (specifically, a likelihood based on normal error assumptions for LAI and C-biomass) to measure how well results from subsequent simulations match these simulated data. We subsequently ran two sensitivity analyses to test for the influence (sensitivity) of 55 LPJ-GUESS parameters (plus two additional parameters for the standard deviations of the normal errors) on the model fit, expressed by the likelihood. Minimum and maximum values for each parameter were derived from the literature (Pappas et al., 2013; Wramneby et al., 2008; Zaehle et al., 2005).

The first analysis performed was a local One-at-A-Time (OAT) sensitivity analysis, which we programmed by hand, using the parallel

option of `rLPJGUESS` with 13 cores (Appendix C, Fig. C1). As a second analysis, we ran a global Morris sensitivity analysis (Campolongo et al., 2007) (Appendix C, Figs. C2 and C3), implemented in the R package `sensitivity` (Pujol et al., 2016). As `rLPJGUESS` allows parallelization of most of the implemented sensitivity functions, we made use of this option. This allowed us to screen a large number of species-specific parameters, while other studies focused on a lower number of parameters at the PFT level, in a relatively short amount of time (~5 min for the OAT analysis, ~15 min for the Morris analysis). The results from our local and global sensitivity analyses (Fig. 3) were in agreement with findings of previous studies, in the sense that the parameters that we identified as particularly important were highlighted (when tested) in previous studies as well (Pappas et al., 2013; Wramneby et al., 2008; Zaehle et al., 2005).
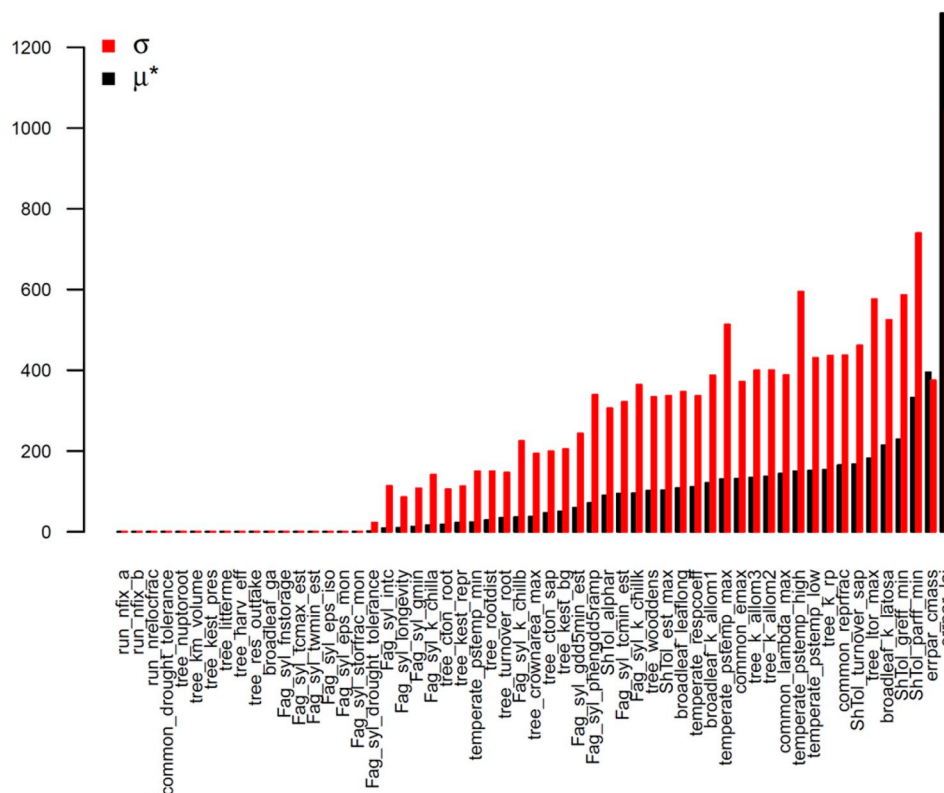

Fig. 3. Results from the Morris sensitivity analysis. The 57 parameters are listed on the x-axis. A high μ* indicates a factor with an important overall influence on model output; a high σ indicates either a factor interacting with other factors or a factor whose effects are non-linear.

3.3. Bayesian calibration

Based on the sensitivity analysis, we selected the 16 most influential parameters and ran a full Bayesian calibration using the differential evolution (DEzs, ter Braak, 2006) MCMC algorithm from the `BayesianTools` R package (Hartig et al., 2017), with 3 chains and 10^5 iterations (Appendix D). The total runtime for this full calibration (unparallelized) was ~118 h (~5 days). As expected for a successful calibration, the model converged on parameter values close to the values used to generate the reference data (Gelman-Rubin's test of convergence: 1.05), with well-defined posterior distributions (Fig. 4a). Predictions from the calibrated model matched the reference data well with small uncertainties (Fig. 4b and c).

`HoltWinter()` and `predict.HoltWinter()` functions (included in the R base package `stats`) to construct hypothetical climate forcing data for the period 2013–2062. Using the calibrated model version from the previous step, we simulated the 10 scenarios in parallel on 10 cores (Fig. 5), with a total runtime of ~5 s. The trends of productivity resulting from these simulations are in accordance with Ahlström et al. (2012), who studied productivity in response to climate change with LPJ-GUESS at this latitude for the time period 1950–2005.

4. Conclusions

While complex system models are widely used in ecology and earth system sciences, they are so far rarely coupled to advanced statistical
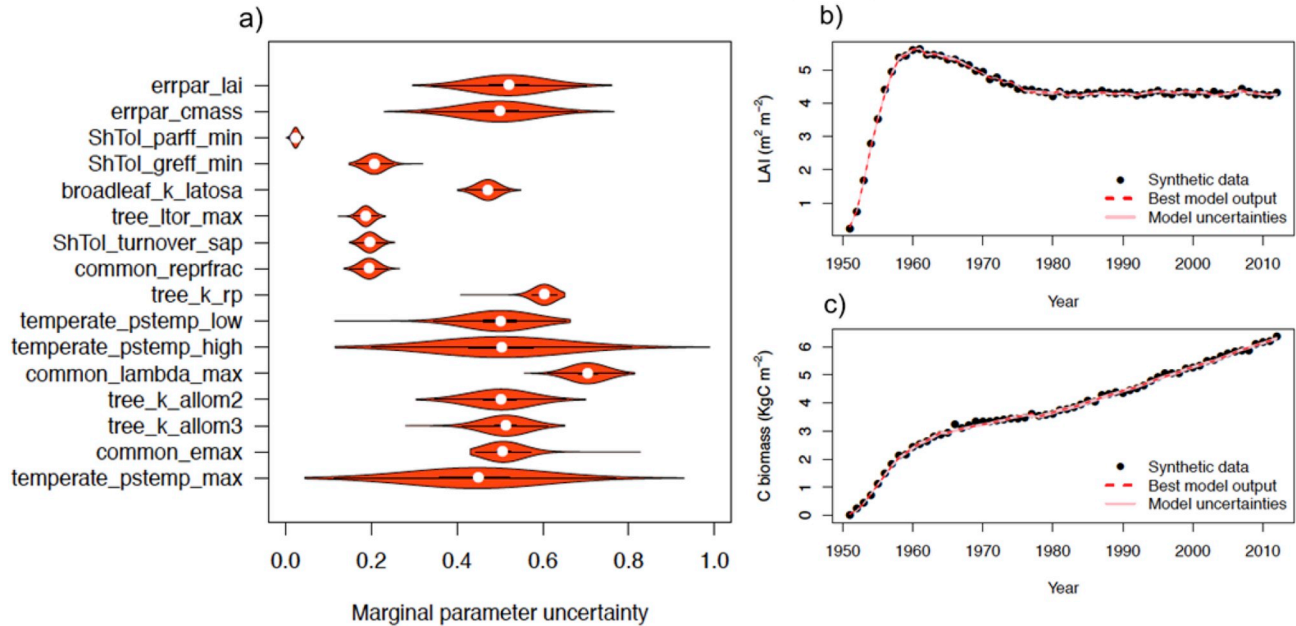

Fig. 4. Results from the calibrated LPJ-GUESS model. a) Marginal posterior distributions of the calibrated parameters, scaled to minimum and maximum values of the prior. White dots indicate the median of the distribution. b) Model/reference data comparison for LAI. c) Model/reference data comparison for C biomass. For b) and c), the best model results were calculated from the Maximum A-Posteriori probability (MAP) parameter estimate.

3.4. Multiple climate scenarios

As a fourth and last analysis step, we set up a simulation with multiple climate scenarios (Appendix E, note that multiple locations could be simulated in the same way). We created 10 time-series, based on the original weather data, with increasing temperature (from +0.5 °C to +5 °C) and precipitations (from +5% to +70%), using the

algorithms for model analysis. rLPJGUESS addresses this issue for the LPJ-GUESS model, allowing users to access many important LPJ-GUESS features from R, with a minimum amount of coding, and advanced parallelization options. In our case study, we highlighted four possible analyses that could be run through rLPJGUESS, but there are many further options. A particular interesting application would be to experiment with model emulations (Fer et al., 2018).
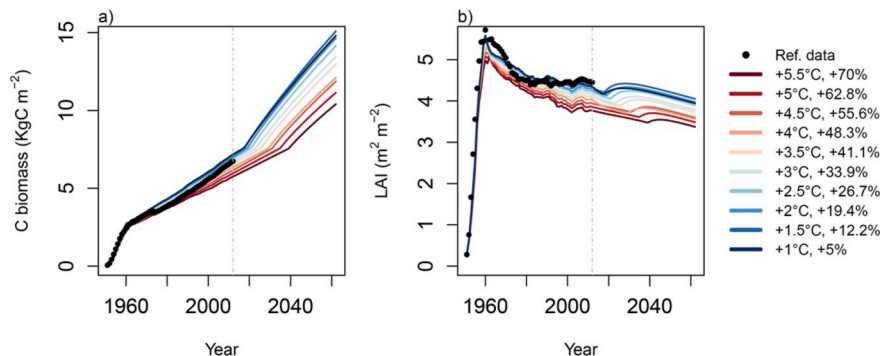

Fig. 5. Model simulations for beech, showing a) carbon biomass, and b) LAI, for multiple climate scenarios. The dotted vertical line is drawn at year 2012, separating past and future projections.

By allowing to investigate LPJ-GUESS in a more systematic and automatized way, `rLPJGUESS` will facilitate new insights about vegetation dynamics as well as a much-needed better integration between the model and the various data streams (see also parallel efforts, e.g. Dietze, 2017). As such, we believe that our package adds significant value to LPJ-GUESS as a tool to investigate the vegetation, carbon, nitrogen and water dynamics of ecosystems over time and to identify new possible threats, especially related to climate change. Moreover, `rLPJGUESS` could be easily extended or adopted to other DGVMs or even other model classes (i.e. climatic or hydrological), in order to improve their comparability and their coupling with unifying existing statistical frameworks (LeBauer et al., 2013; Voosen, 2016).

Conflicts of interest

All authors declare no conflict of interests.

Acknowledgments

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2018.09.004.

References

Ahlström, A., Raupach, M.R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J.G., Friedlingstein, P., Jain, A.K., Kato, E., Poulter, B., Sitch, S., Stocker, B.D., Viovy, N., Wang, Y.P., Wiltshire, A., Zaehle, S., Zeng, N., 2015. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science 80 (348), 895–899.

Ahlström, A., Schurgers, G., Arneth, A., Smith, B., 2012. Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. Environ. Res. Lett. 7, 044008. https://doi.org/10.1088/1748-9326/7/4/044008. (9pp).

Bagnara, M., Sottocornola, M., Cescatti, A., Minerbi, S., Montagnani, L., Gianelle, D., Magnani, F., 2015. Bayesian optimization of a light use efficiency model for the estimation of daily gross primary productivity in a range of Italian forest ecosystems. Ecol. Model. 306. https://doi.org/10.1016/j.ecolmodel.2014.09.021.

Bagnara, M., Van Oijen, M., Cameron, D., Gianelle, D., Magnani, F., Sottocornola, M., 2018. Bayesian calibration of simple forest models with multiplicative mathematical structure: a case study with two Light Use Efficiency Models in an alpine forest. Ecol.

Model. 371, 90–100. https://doi.org/10.1016/j.ecolmodel.2018.01.014.

Bugmann, H., 2001. A review of forest gap models. Climatic Change 51, 259–305.

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environ. Model. Software 22, 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004.

Dietze, M.C., 2017. Ecological Forecasting. Princeton University Press, Princeton, NJ.

Dietze, M.C., Serbin, S.P., Davidson, C., Desai, A.R., Feng, X., Kelly, R., Kooper, R., LeBauer, D., Mantooth, J., McHenry, K., Wang, D., 2014. A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. J. Geophys. Res. Biogeosci. 119, 286–300. https://doi.org/10.1002/2013JG002392.

Fer, I., Kelly, R., Moorcroft, P.R., Richardson, A.D., Cowdery, E.M., Dietze, M.C., 2018. Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. Biogeosci. Discuss. 1–30. https://doi.org/10.5194/bg-2018-96.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.D., Weisser, W.W., 2010. Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. Basic Appl. Ecol. 11, 473–485.

Granier, A., Bréda, N., Longdoz, B., Gross, P., Ngao, J., 2008. Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France. Ann. For. Sci. 65, 704. https://doi.org/10.1051/forest:2008052.

Hartig, F., Calabrese, J.M., Reineking, B., Wiegand, T., Huth, A., 2011. Statistical inference for stochastic simulation models - theory and application. Ecol. Lett. 14, 816–827.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S., Huth, A., 2012. Connecting dynamic vegetation models to data – an inverse perspective. J. Biogeogr. 39, 2240–2252. https://doi.org/10.1111/j.1365-2699.2012.02745.x.

Hartig, F., Minunno, F., Paul, S., 2017. BayesianTools: General-purpose MCMC and SMC Samplers and Tools for Bayesian Statistics.

Hickler, T., Vohland, K., Feehan, J., Miller, P.A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T.R., Cramer, W., Kühn, I., Sykes, M.T., 2012. Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. Global Ecol. Biogeogr. 21, 50–63. https://doi.org/10.1111/j.1466-8238.2010.00613.x.

LeBauer, D.S., Wang, D., Richter, K.T., Davidson, C.C., Dietze, M.C., 2013. Facilitating feedbacks between field measurements and ecosystem models. Ecol. Monogr. 83, 133–154. https://doi.org/10.1890/12-0137.1.

Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., Zhang, L., 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. Ecol. Appl. 19, 571–574.

Pappas, C., Fatichi, S., Leuzinger, S., Wolf, A., Burlando, P., 2013. Sensitivity analysis of a process-based ecosystem model: pinpointing parameterization and structural issues. J. Geophys. Res. Biogeosci. 118, 505–528. https://doi.org/10.1002/jgrg.20035.

Prentice, I.C., Bondeau, A., Cramer, W., Harrison, S.P., Hickler, T., Lucht, W., Sitch, S., Smith, B., Sykes, M.T., 2007. Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. In: Canadell, J.G., Pataki, D.E., Pitelka, L.F. (Eds.), Terrestrial Ecosystems in a Changing World. Springer-Verlag, New York, pp. 175–192.

Pujol, G., Iooss, B., with contributions from Khalid Boumhaout, A.J,Veiga, S. Da, Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L. Le, Lemaitre, P., Ramos, B., Touati, T., Weber, F., 2016. Sensitivity: Global Sensitivity Analysis of Model Outputs.

R Core Team, 2015. R: a Language and Environment for Statistical Computing.

Saltelli, A., Chan, K., Scott, E.M., 2000. Sensitivity Analysis. Wiley, Chichester, U.K.

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S., 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biol. 9, 161–185. https://doi.org/10.1046/j.1365-2486.2003.00569.x.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., Zaehle, S., 2014. Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences 11, 2027–2054.

ter Braak, C.J.F., 2006. A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. Stat. Comput. 16, 239–249. https://doi.org/10.1007/s11222-006-8769-1.

van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol. 25, 915–927. https://doi.org/10.1093/treephys/25.7.915.

Voosen, P., 2016. Climate scientists open up their black boxes to scrutiny. Science 80 (354), 401–402.

Wramneby, A., Smith, B., Zaehle, S., Sykes, M.T., 2008. Parameter uncertainties in the modelling of vegetation dynamics-Effects on tree community structure and ecosystem functioning in European forest biomes. Ecol. Model. 216, 277–290.

Zaehle, S., Sitch, S., Smith, B., Hatterman, F., 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochem. Cycles 19, 1–16.